

Customer No. 20350

TOWNSEND and TOWNSEND and CREW LLP  
Two Embarcadero Center, 8<sup>th</sup> Floor  
San Francisco, California 94111-3834  
(415) 576-0200

ASSISTANT COMMISSIONER FOR PATENTS  
BOX PATENT APPLICATION  
Washington, D.C. 20231

Attorney Docket No. 018547-018570US

"Express Mail" Label No. EM035146244US

Date of Deposit: April 2, 1999

I hereby certify that this is being deposited with the United States  
Postal Service "Express Mail Post Office to Addressee" service  
under 37 CFR 1.10 on the date indicated above, addressed to:

Assistant Commissioner for Patents  
Washington, D.C. 20231

By: *Kenia M. Van der ...*

Inventor(s)/Applicant Identifier: Thomas Gingeras and Jorg Dreukow

For: MYCOBACTERIAL RpoB SEQUENCES

[X] This application claims priority from each of the following Application Nos./filing dates:

60/080,616, filed April 3, 1998, the disclosure of which is incorporated by reference.

[X] Please amend this application by adding the following before the first sentence: "This application claims the benefit of U.S.  
Provisional Application No. 60/080,616, filed April 3, 1998, the disclosure of which is incorporated by reference."

Enclosed are:

[X] 20 pages of Specification, 3 pages of Claims, 1 page of Abstract, 56 sheets of Drawings.

[X] Unsigned Declaration and Power of Attorney;

[X] Return Postcard.

**In view of the Unsigned Declaration as filed with this application and pursuant to 37 CFR §1.53(f),  
Applicant requests deferral of the filing fee until submission of the Missing Parts of Application.**

DO NOT CHARGE THE FILING FEE AT THIS TIME.

Telephone:  
(415) 576-0200

Facsimile:  
(415) 576-0300

*Joe Liebeschuetz*  
Joe Liebeschuetz  
Reg No.: 37,505  
Attorneys for Applicant

PATENT APPLICATION

**MYCOBACTERIAL *RpoB* SEQUENCES**

Inventor: Thomas Gingeras  
Jorg Drenkow

Assignee: AFFYMETRIX, INC.  
3380 Central Expressway  
Santa Clara, California 95051  
a Corporation of California

Entity: Large

TOWNSEND and TOWNSEND and CREW LLP  
Two Embarcadero Center, 8th Floor  
San Francisco, California 94111-3834  
(415) 326-2400

662040" 90E52260

5

**MYCOBACTERIAL *RpoB* SEQUENCES**

## STATEMENT OF GOVERNMENT INTEREST

The work described in this application was supported in part by grant number 1R43A140400 by the NIAID. The Government may have certain rights in this invention.

10

## CROSS-REFERENCE TO RELATED APPLICATIONS

15

This application derives priority from USSN 60/080,616, filed April 3, 1999, and incorporated by reference. Applications USSN 08/797,812, filed February 7, 1997, USSN 60/011,339, filed 08 Feb. 1996; 60/012,631, filed 01 March 1996; 08/629,031, filed 08 April 1996; and 60/017,765, filed 15 May 1996 are directed to related subject matter. These applications are specifically incorporated by reference in their entirety for all purposes.

20

## BACKGROUND OF THE INVENTION

Field of the Invention

25

This invention is directed to polymorphisms in *rpoB* genes of mycobacteria and use of the same in the identification and characterization of microorganisms.

Background of the Invention

30

Multidrug resistance and human immunodeficiency virus (HIV-1) infections are factors which have had a profound impact on the tuberculosis problem. An increase in the frequency of *Mycobacterium tuberculosis* strains resistant to one or more anti-mycobacterial agents has been reported, Block, et al., (1994) JAMA **271**:665-671. Immunocompromised HIV-1 infected patients not infected with *M. tuberculosis* are

35

662040 " 90E33260

frequently infected with *M. avium* complex (MAC) or *M. avium*-*M. intracellulare* (MAI) complex. These mycobacteria species are often resistant to the drugs used to treat *M. tuberculosis*. These factors have re-emphasized the importance for the accurate determination of drug sensitivities and mycobacteria species identification.

In HIV-1 infected patients, the correct diagnosis of the mycobacterial disease is essential since treatment of *M. tuberculosis* infections differs from that called for by other mycobacteria infections, Hoffner, S.E. (1994) Eur. J. Clin. Microbiol. Inf. Dis. **13**:937-941. Non-tuberculosis mycobacteria commonly associated with HIV-1 infections include *M. kansasii*, *M. xenopi*, *M. fortuitum*, *M. avium* and *M. intracellulare*, Wolinsky, E., (1992) Clin. Infect. Dis. **15**:1-12, Shafer, R.W. and Sierra, M.F. 1992 Clin. Infect. Dis. **15**:161-162. Additionally, 13% of new cases (HIV-1 infected and non-infected) of *M. tuberculosis* are resistant to one of the primary anti-tuberculosis drugs (isoniazid [INH], rifampin [RIF], streptomycin [STR], ethambutol [EMB] and pyrazinamide [PZA] and 3.2% are resistant to both RIF and INH, Block, et al., JAMA **271**:665-671, (1994). Consequently, mycobacterial species identification and the determination of drug resistance have become central concerns during the diagnosis of mycobacterial diseases.

Methods used to detect, and to identify *Mycobacterium* species vary considerably. For detection of *Mycobacterium tuberculosis*, microscopic examination of acid-fast stained smears and cultures are still the methods of choice in most microbiological clinical laboratories. However, culture of clinical samples is hampered by the slow growth of mycobacteria. A mean time of four weeks is required before sufficient growth is obtained to enable detection and possible identification. Recently, two more rapid methods for culture have been developed involving a radiometric, Stager, C.E. et al., (1991) J. Clin. Microbiol. **29**:154-157, and a

biphasic (broth/agar) system Sewell, et al., (1993) J. Clin. Microbiol. **29**:2689-2472. Once grown, cultured mycobacteria can be analyzed by lipid composition, the use of species specific antibodies, species specific DNA or RNA probes and PCR-based sequence analysis of 16S rRNA gene (Schirm, et al. (1995) J. Clin. Microbiol. **33**:3221-3224; Kox, et al. (1995) J. Clin. Microbiol. **33**:3225-3233) and IS6110 specific repetitive sequence analysis (For a review see, e.g., Small et al., P.M. and van Embden, J.D.A. (1994) Am. Society for Microbiology, pp. 569-582). The analysis of 16S rRNA sequences (RNA and DNA) has been the most informative molecular approach to identify *Mycobacteria* species (Jonas, et al., J. Clin. Microbiol. **31**:2410-2416 (1993)). However, to obtain drug sensitivity information for the same isolate, additional protocols (culture) or alternative gene analysis is necessary.

To determine drug sensitivity information, culture methods are still the protocols of choice. *Mycobacteria* are judged to be resistant to particular drugs by use of either the standard proportional plate method or minimal inhibitory concentration (MIC) method. However, given the inherent lengthy times required by culture methods, approaches to determine drug sensitivity based on molecular genetics have been recently developed.

Because resistance to RIF in *E. coli* strains was observed to arise as a result of mutations in the *rpoB* gene, Telenti, et al., id., identified a 69 base pair (bp) region of the *M. tuberculosis rpoB* gene as the locus where RIF resistant mutations were focused. Kapur, et al., (1995) Arch. Pathol. Lab. Med. **119**:131-138, identified additional novel mutations in the *M. tuberculosis rpoB* gene which extended this core region to 81 bp. In a detailed review on antimicrobial agent resistance in mycobacteria, Musser (Clin. Microbiol. Rev., **8**:496-514 (1995)), summarized all the characterized mutations and their relative frequency of occurrence in this 81 bp region of *rpoB*. Missense mutations comprise 88% of all known

mutations while insertions (3 or 6 bp) and deletions (3, 6 and 9 bp) account for 4% and 8% of the remaining mutations, respectively. Approximately 90% of all RIF resistant tuberculosis isolates have been shown to have mutations in this 81 bp region. The remaining 10% are thought possibly to involve genes other than *rpoB*.

For the above reasons, it would be desirable to have simpler methods which identify and characterize microorganisms, such as *Mycobacteria*, both at the phenotypic and genotypic level. This invention fulfills that and related needs.

#### SUMMARY OF THE INVENTION

In one aspect, the invention provides isolated nucleic acids comprising at least 25, 50, 75, 100, or 200 contiguous bases from an *rpoB* sequence shown in Table 1. Some nucleic acid comprise a complete sequence shown in Table 1.

The invention further provides a set of probes perfectly complementary to and spanning such nucleic acids, preferably spanning one of the complete sequences shown in Table 1.

The invention further provides methods of classifying mycobacteria. Some such methods entail providing a sample comprising a mycobacterial *rpoB* target nucleic acid from a mycobacteria, determining the sequence of a segment of at least 50 contiguous bases from the target nucleic acid; comparing the determined sequence to at least one sequence shown in Table 1; and classifying the mycobacteria from the extent of similarity of the compared sequences. Preferably, at least 100 or 200 contiguous bases are determined from the target nucleic acid. Preferably, the determined sequence is compared with a plurality of sequences from Table 1, for example, 10, 20, 50 or all of the sequence from Table 1.

In other methods of classification, the identity of one or more bases in the target sequence at one or more

positions corresponding to one or more of the highlighted positions in a sequence shown in Table 1 is determined. The identity of the one or more bases characterizing the species of mycobacteria that is present in the sample. In some methods, the identity of at least 10 bases in the target nucleic acid at positions corresponding to highlighted positions in a sequence shown in Table 1 is determined. In some methods, the identity of at least 20 bases in the target sequence at highlighted positions shown in Table 1 are identified. In some methods, at least 20 determined bases are compared with 20 bases occupying corresponding positions in each of at least ten sequences from Table 1.

In another aspect, the invention provides sequence-specific polynucleotide probes or primers that hybridizes to a segment of a mycobacterial rpoB sequence shown in Table 1 or its complement without hybridizing to the *M. tuberculosis* sequence designated ATCC9-Mtb in Table 1 or its complement, the segment including a highlighted nucleotide position shown in Table 1. In some such probes, a central position of the probe aligns with a highlighted nucleotide position shown in Tables 1. In some such primers, the 3' end of the primer aligns with a highlighted nucleotide position shown in Table 1. Some probes and primers are between 10 and 50 bases long.

In another aspect, the invention provides a computer-readable storage medium for storing data for access by an application program being executed on a data processing system. Such a system comprises a data structure stored in the computer-readable storage medium. The data structure includes information resident in a database used by the application program and includes a plurality of records, each record comprising information identifying a polymorphism or sequence shown in Table 1. Some records have a field identifying a base occupying a polymorphic site and a field identifying location of the polymorphic site. Some records record a contiguous segment of at least 50, 100, or 200 bases

from an rpoB sequence shown in Table 1. Some storage medium comprise at least ten records each recording a contiguous segment of at least 50 bases from at least ten rpoB sequences shown in Table 1.

#### BRIEF DESCRIPTION OF THE DRAWINGS

Fig. 1: Computer that may be utilized to execute software embodiments of the present invention.

Fig. 2: A system block diagram of a typical computer system that may be used to execute software embodiments of the invention.

#### DEFINITIONS

A polynucleotide can be DNA or RNA, and single- or double-stranded. Polynucleotide can be naturally occurring or synthetic, and can be of any length. Preferred polynucleotide probes of the invention include contiguous segments of DNA, or their complements including any of the highlighted bases shown in Table 1. The segments are usually between 5 and 100 bases, and often between 5-10, 5-20, 10-20, 10-50, 20-50 or 20-100 bases. The highlighted site can occur within any position of the segment. Preferred polynucleotide probes are capable of binding in a base-specific manner to a complementary strand of nucleic acid. Such probes include peptide nucleic acids, as described in Nielsen et al., *Science* 254, 1497-1500 (1991), and probes having nonnaturally occurring bases.

The term primer refers to a single-stranded polynucleotide capable of acting as a point of initiation of template-directed DNA synthesis under appropriate conditions (i.e., in the presence of four different nucleoside triphosphates and an agent for polymerization, such as, DNA or RNA polymerase or reverse transcriptase) in an appropriate buffer and at a suitable temperature. The appropriate length of a primer depends on the intended use of the primer but typically ranges from 15 to 30 nucleotides. Short primer



molecules generally require cooler temperatures to form sufficiently stable hybrid complexes with the template. The term primer site refers to the area of the target DNA to which a primer hybridizes. The term primer pair means a set of primers including a 5' upstream primer that hybridizes with the 5' end of the DNA sequence to be amplified and a 3', downstream primer that hybridizes with the complement of the 3' end of the sequence to be amplified.

A cDNA or cRNA is derived from an RNA if it produced by a process in which the RNA serves as a template for production of the cDNA or cRNA.

Hybridizations are usually performed under stringent conditions, for example, at a salt concentration of no more than 1 M and a temperature of at least 25°C. For example, conditions of 5 x SSPE (750 mM NaCl, 50 mM Na Phosphate, 5 mM EDTA, pH 7.4) and a temperature of 25-30°C are suitable for allele-specific probe hybridizations.

An isolated nucleic acid means an object species invention that is the predominant species present (*i.e.*, on a molar basis it is more abundant than any other individual species in the composition). Preferably, an isolated nucleic acid comprises at least about 50, 80 or 90 percent (on a molar basis) of all macromolecular species present. Most preferably, the object species is purified to essential homogeneity (contaminant species cannot be detected in the composition by conventional detection methods).

For sequence comparison and homology determination, typically one sequence acts as a reference sequence to which test sequences are compared. When using a sequence comparison algorithm, test and reference sequences are input into a computer, subsequence coordinates are designated, if necessary, and sequence algorithm program parameters are designated. The sequence comparison algorithm then calculates the percent sequence identity for the test sequence(s) relative to the reference sequence, based on the designated program parameters.

Optimal alignment of sequences for comparison can be conducted, e.g., by the local homology algorithm of Smith & Waterman, *Adv. Appl. Math.* 2:482 (1981), by the homology alignment algorithm of Needleman & Wunsch, *J. Mol. Biol.* 48:443 (1970), by the search for similarity method of Pearson & Lipman, *Proc. Nat'l. Acad. Sci. USA* 85:2444 (1988), by computerized implementations of these algorithms (GAP, BESTFIT, FASTA, and TFASTA in the Wisconsin Genetics Software Package, Genetics Computer Group, 575 Science Dr., Madison, WI), or by visual inspection (see generally, Ausubel *et al.*, *infra*).

One example of algorithm that is suitable for determining percent sequence identity and sequence similarity is the BLAST algorithm, which is described in Altschul *et al.*, *J. Mol. Biol.* 215:403-410 (1990). Software for performing BLAST analyses is publicly available through the National Center for Biotechnology Information (<http://www.ncbi.nlm.nih.gov/>). This algorithm involves first identifying high scoring sequence pairs (HSPs) by identifying short words of length W in the query sequence, which either match or satisfy some positive-valued threshold score T when aligned with a word of the same length in a database sequence. T is referred to as the neighborhood word score threshold (Altschul *et al.*, *supra*). These initial neighborhood word hits act as seeds for initiating searches to find longer HSPs containing them. The word hits are then extended in both directions along each sequence for as far as the cumulative alignment score can be increased. Cumulative scores are calculated using, for nucleotide sequences, the parameters M (reward score for a pair of matching residues; always > 0) and N (penalty score for mismatching residues; always < 0). For amino acid sequences, a scoring matrix is used to calculate the cumulative score. Extension of the word hits in each direction are halted when: the cumulative alignment score falls off by the quantity X from its maximum achieved value; the cumulative score goes to zero or below, due to the

accumulation of one or more negative-scoring residue alignments; or the end of either sequence is reached. The BLAST algorithm parameters W, T, and X determine the sensitivity and speed of the alignment. The BLASTN program (for nucleotide sequences) uses as defaults a wordlength (W) of 11, an expectation (E) of 10, a cutoff of 100, M=5, N=-4, and a comparison of both strands. For amino acid sequences, the BLASTP program uses as defaults a wordlength (W) of 3, an expectation (E) of 10, and the BLOSUM62 scoring matrix (see Henikoff & Henikoff (1989) *Proc. Natl. Acad. Sci. USA* 89:10915).

In addition to calculating percent sequence identity, the BLAST algorithm also performs a statistical analysis of the similarity between two sequences (see, e.g., Karlin & Altschul (1993) *Proc. Nat'l. Acad. Sci. USA* 90:5873-5787). One measure of similarity provided by the BLAST algorithm is the smallest sum probability (P(N)), which provides an indication of the probability by which a match between two nucleotide or amino acid sequences would occur by chance. For example, a nucleic acid is considered similar to a reference sequence if the smallest sum probability in a comparison of the test nucleic acid to the reference nucleic acid is less than about 0.1, more preferably less than about 0.01, and most preferably less than about 0.001.

The term "target nucleic acid" refers to a nucleic acid (often derived from a biological sample), to which the probe nucleic acid is designed to specifically hybridize. It is the presence or expression level of the target nucleic acid that is to be detected or quantified. The target nucleic acid has a sequence that is complementary to the nucleic acid sequence of the corresponding probe directed to the target. The term target nucleic acid may refer to the specific subsequence of a larger nucleic acid to which the probe is directed or to the overall sequence (e.g. gene or mRNA) whose expression level it is desired to detect. The difference in usage will be apparent from context.

"Subsequence" refers to a sequence of nucleic acids that comprise a part of a longer sequence of nucleic acids.

5

## DETAILED DESCRIPTION

### I) Mycobacterial Sequences of rpoB Genes

10

15

20

25

Table 1 shows a comparison of a substantial collection of mycobacterial strains of an about 700-nucleotide conserved region of an rpoB gene. The first sequence, designated as a reference sequence, is from *M. tuberculosis*. Nucleotides are numbered consecutively starting from the first nucleotide of the reference sequences. Other sequences are from other strains of mycobacteria. For example, the sequences designated ATCC-av, M29, M30...M104 are from *M. avium*. Sequences designated from ATT-chelnew, M11, M13, and M17 are from *M. chelonae*. Sequences designated ATCC-for, M53, M55, M56, and M74 are from *M. fortuitum*, and so fourth. Complete correspondence between strain designations and strain types is shown in Table 2. Nucleotides in a mycobacterial sequence are accorded the same number as the corresponding position of the reference sequence when the two are maximally aligned. Differences between a sequence and the reference sequences are shown in highlighted type. Many of the highlighted positions are common to all tested members of a species. Other highlighted positions vary among different isolates in a species. Both types of variation can be useful in speciation analysis.

### II. Analysis of Species Variations

30

#### A. Preparation of Samples

35

An rpoB sequence is isolated from a sample of an unknown mycobacteria being tested. Nucleic acids can be isolated from mycobacteria by standard methods as described in WO 97/29212 (incorporated by reference in its entirety for all purposes). The rpoB sequences to be analyzed can then be isolated and amplified by means of PCR. See generally PCR

*Technology: Principles and Applications for DNA Amplification* (ed. H.A. Erlich, Freeman Press, NY, NY, 1992); *PCR Protocols: A Guide to Methods and Applications* (eds. Innis, et al., Academic Press, San Diego, CA, 1990); Mattila et al., *Nucleic Acids Res.* 19, 4967 (1991); Eckert et al., *PCR Methods and Applications* 1, 17 (1991); *PCR* (eds. McPherson et al., IRL Press, Oxford); and U.S. Patent 4,683,202 (each of which is incorporated by reference for all purposes). Primers for PCR preferably flank the regions of interest rpoB genes, although primers to internal sites can be used if it is intended to analyze only certain sites of potential species variation. Exemplary primers are described in WO 97/29212. If necessary, additional sequences flanking the sequences shown in Table 1 can be determined using probes based on the sequences in Table 1 to isolate full-length rpoB sequences from the appropriate mycobacterial species.

B. Detection of Species-Specific Variations in Target DNA

1. Sequence-Specific Probes

The design and use of sequence-specific probes for analyzing polymorphisms is described by e.g., Saiki et al., *Nature* 324, 163-166 (1986); Dattagupta, EP 235,726, Saiki, WO 89/11548. Sequence-specific probes can be designed that hybridize to a segment of target DNA in one isolate of mycobacteria that do not isolate to a corresponding isolate in another due to the presence of allelic or species variations in the respective segments from the two sequences. Hybridization conditions should be sufficiently stringent that there is a significant difference in hybridization intensity between alleles, and preferably an essentially binary response, whereby a probe hybridizes to only one of the sequences. Some probes are designed to hybridize to a segment of target DNA such that the site of potential sequence variation aligns with a central position (e.g., in a 15 mer at the 7 position; in a 16 mer, at either the 8 or 9 position) of

the probe. This design of probe achieves good discrimination in hybridization between different allelic and species variants.

Sequence-specific probes are often used in pairs, one member of a pair showing a perfect match to a reference form of a target sequence and the other member showing a perfect match to a variant form. Several pairs of probes can then be immobilized on the same support for simultaneous analysis of multiple potential variations within the same target sequence.

## 2. Tiling Arrays

The bases occupying sites of potential variation can also be identified by hybridization to nucleic acid arrays, some example of which are described by WO 95/11995 (incorporated by reference in its entirety for all purposes). Such arrays contain a series of overlapping probes spanning a reference sequence. Any of the *rpoB* sequences shown in Table 1, or contiguous segments of, for example, at least 25, 50, 100 or 200 bases thereof, can serve as a reference sequence. WO 95/11995 also describes subarrays that are optimized for detection of a variant forms of a precharacterized polymorphism. Such a subarray contains probes designed to be complementary to a second reference sequence, which is a variant of the first reference sequence. The inclusion of a second group (or further groups) can be particular useful for analyzing short subsequences of the primary reference sequence in which multiple mutations are expected to occur within a short distance commensurate with the length of the probes (*i.e.*, two or more mutations within 9 to 21 bases).

## 3. Sequence-Specific Primers

A sequence-specific primer hybridizes to a site on target DNA overlapping a polymorphism and only primes amplification of a variant form to which the primer exhibits perfect complementarity. See Gibbs, *Nucleic Acid Res.* 17,

2427-2448 (1989). This primer is used in conjunction with a second primer which hybridizes at a distal site.

Amplification proceeds from the two primers leading to a detectable product signifying the particular variant form is present. A control is usually performed with a second pair of primers, one of which shows a single base mismatch at the site of variation and the other of which exhibits perfect complementarily to a distal site. The single-base mismatch prevents amplification and no detectable product is formed.

The method works best when the mismatch is included in the 3'-most position of the primer aligned with the with the point of variation because this position is most destabilizing to elongation from the primer. See, e.g., WO 93/22456.

#### 4. Direct-Sequencing

The direct analysis of mycobacterial sequences can be accomplished using either the dideoxy chain termination method or the Maxam Gilbert method (see Sambrook et al., *Molecular Cloning, A Laboratory Manual* (2nd Ed., CSHP, New York 1989); Zyskind et al., *Recombinant DNA Laboratory Manual*, (Acad. Press, 1988)).

### III. Methods of Use

The sequences and polymorphisms shown in Table 1 are useful for identifying the presence of mycobacteria in samples, and optionally, classifying the mycobacteria. The sample can be obtained from a patient or from a biological source, such as a food product.

The sequences shown in Table 1 can be used for design of sequence-specific probes or primers encompassing polymorphic sites as described above. These probes or primers can then be used to determine the base occupying a corresponding position in an rpoB sequence from an isolate in a sample under test. A base in one sequence corresponds with a base in another when the two bases occupy the same position

when the two sequences are maximally aligned by one of the criteria described in Definitions.

Alternatively, the sequences shown in Table 1 can be used for design of tiling arrays in which one or more of the sequences serves as a reference sequence. At least one set of overlapping probes is designed spanning a segment of the reference sequence, as described in WO95/11995 or EP 717,113. Target sequences from samples under test can be hybridized to such arrays, optionally in combination with controls of known rpoB sequences. The hybridization pattern of a target sequence to such an array can be analyzed to determine the identity of bases at which the target sequence differs from the reference sequence, as described in WO 95/11995.

One or more of the above methods, or direct sequencing, can be used to identify the base occupying at least one and usually several (e.g., 5, 10, 15, 25, 50 or 100) sites of potential variation between the 16S RNA and/or rpoB gene in an unknown mycobacteria relative to bases occupying corresponding sites in one or more known strains of mycobacteria, such as those shown in Tables 1. This analysis results in a profile of bases occupying particular sites that characterizes the mycobacterial strain under test. The profile is compared with the corresponding profiles of different mycobacterial isolates shown in e.g., Tables 1. In general, the unknown mycobacterium isolate is characterized as being from the same mycobacterial species as the precharacterized isolate with which it shares the greatest similarity in base profile.

In some methods, the sequence of a contiguous segment of the rpoB target nucleic acid is determined in a sample under test for comparison with one or more of the sequences shown in Table 1. The mycobacteria is classified by the extent of similarity. For example, if a target nucleic acid shows greater sequence identity to rpoB sequences from one species than any other, the sample from which the target



was obtained is typically classified as arising from that species.

Alternatively, an array of tiled probes based on a reference sequence shown in Table 1 can be used for identifying and characterizing mycobacterial sequences based on comparison of hybridization patterns. Such an array is hybridized to a 16S RNA or rpoB target sequence from a sample, and the hybridization pattern compared with the hybridization pattern of one or more control sequences. The hybridization patterns of control sequences can be historic controls, stored, for example, in a computer database, or can be contemporaneous controls performed at or near the same time as the hybridization to the target sequence. Optionally, hybridization of target and reference sequence can be performed simultaneously using different labels.

Method of classifying unknown mycobacterial isolate by matching the hybridization pattern of a target sequence with those of control sequences from characterized species are described in more detail in WO 97/29212 (incorporated by reference in its entirety for all purposes). In an idealized case, the detection of a particular hybridization pattern in an isolate characterizes that isolate as belonging to a particular species. This can occur when the hybridization pattern detected in the isolate is uniquely associated with a specific species. More frequently however, such an unique one-to-one correspondence is not present. Instead, the hybridization pattern observed in an isolate does not bear a unique correspondence with a previously characterized species. However, the hybridization pattern detected is associated with a probability of the organism being screened belonging to a particular species (or not) or carrying a particular phenotypic trait (or not). As a result, analysis of an increasing number of polymorphic sites in an isolate, allows one to classify the isolated with an increasing level of confidence. Algorithms can be used to derive such composite probabilities from the comparison of multiple polymorphic

forms between an isolate and references. Typically, the mathematical algorithm makes a call of the identity of the species and assign a confidence level to that call. One can determine the confidence level (>90%, >95% etc.) that one desires and the algorithm will analyze the hybridization pattern and either provide an identification or not. Occasionally, the call is that the sample may be one of two, three or more species, in which case a specific identification is not be possible. However, one of the strengths of this technique is that the rapid screening made possible by the chip-based hybridization allows one to continuously expand a database of patterns ultimately to enable the identification of species previously unidentifiable due to lack of sufficient information.

#### IV. Modified Polypeptides and Gene Sequences

The invention further provides variant forms of nucleic acids and corresponding proteins. The nucleic acids comprise one of the sequences described in Tables 1. Some nucleic acid encode full-length variant forms of proteins. Variant proteins have the prototypical amino acid sequences of encoded by nucleic acid sequence shown in Tables 1 (read so as to be in-frame with the full-length coding sequence of which it is a component).

Variant genes can be expressed in an expression vector in which a variant gene is operably linked to a native or other promoter. Usually, the promoter is a eukaryotic promoter for expression in a mammalian cell. The transcription regulation sequences typically include a heterologous promoter and optionally an enhancer which is recognized by the host. The selection of an appropriate promoter, for example trp, lac, phage promoters, glycolytic enzyme promoters and tRNA promoters, depends on the host selected. Commercially available expression vectors can be used. Vectors can include host-recognized replication

systems, amplifiable genes, selectable markers, host sequences useful for insertion into the host genome, and the like.

The means of introducing the expression construct into a host cell varies depending upon the particular construction and the target host. Suitable means include fusion, conjugation, transfection, transduction, electroporation or injection, as described in Sambrook, *supra*. A wide variety of host cells can be employed for expression of the variant gene, both prokaryotic and eukaryotic. Suitable host cells include bacteria such as *E. coli*, yeast, filamentous fungi, insect cells, mammalian cells, typically immortalized, e.g., mouse, CHO, human and monkey cell lines and derivatives thereof. Preferred host cells are able to process the variant gene product to produce an appropriate mature polypeptide. Processing includes glycosylation, ubiquitination, disulfide bond formation, general post-translational modification, and the like.

The protein may be isolated by conventional means of protein biochemistry and purification to obtain a substantially pure product, i.e., 80, 95 or 99% free of cell component contaminants, as described in Jacoby, *Methods in Enzymology* Volume 104, Academic Press, New York (1984); Scopes, *Protein Purification, Principles and Practice*, 2nd Edition, Springer-Verlag, New York (1987); and Deutscher (ed), *Guide to Protein Purification, Methods in Enzymology*, Vol. 182 (1990). If the protein is secreted, it can be isolated from the supernatant in which the host cell is grown. If not secreted, the protein can be isolated from a lysate of the host cells.

In addition to substantially full-length polypeptides expressed by variant genes, the present invention includes biologically active fragments of the polypeptides, or analogs thereof, including organic molecules which simulate the interactions of the peptides. Biologically active fragments include any portion of the full-length polypeptide which confers a biological function on the variant gene

product, including ligand binding, and antibody binding. Ligand binding includes binding by nucleic acids, proteins or polypeptides, small biologically active molecules, or large cellular structures.

5 Polyclonal and/or monoclonal antibodies that specifically bind to variant gene products but not to corresponding prototypical gene products are also provided. Antibodies can be made by injecting mice or other animals with the variant gene product or synthetic peptide fragments  
10 thereof. Monoclonal antibodies are screened as are described, for example, in Harlow & Lane, *Antibodies, A Laboratory Manual*, Cold Spring Harbor Press, New York (1988); Goding, *Monoclonal antibodies, Principles and Practice* (2d ed.) Academic Press, New York (1986). Monoclonal antibodies are  
15 tested for specific immunoreactivity with a variant gene product and lack of immunoreactivity to the corresponding prototypical gene product. These antibodies are useful in diagnostic assays for detection of the variant form, or as an active ingredient in a pharmaceutical composition.

#### 20 V. Kits

The invention further provides kits comprising at least one sequence-specific probe as described above. Often, the kits contain one or more pairs of sequence-specific probes  
25 hybridizing to different forms of a polymorphism. In some kits, the sequence-specific probes are provided immobilized to a substrate. For example, the same substrate can comprise sequence-specific probes for detecting at least 10, 100 or all of the variations shown in Table 1. Optional additional  
30 components of the kit include, for example, restriction enzymes, reverse-transcriptase or polymerase, the substrate nucleoside triphosphates, means used to label (for example, an avidin-enzyme conjugate and enzyme substrate and chromogen if the label is biotin), and the appropriate buffers for reverse  
35 transcription, PCR, or hybridization reactions. Usually, the kit also contains instructions for carrying out the methods.

## VI. Computer Databases

Fig. 1 illustrates an example of a computer system that can be used to store records relating to polymorphisms of the invention and perform algorithms comparing polymorphic profiles and to classify species. Fig. 32 shows a computer system 100 which includes a monitor 102, screen 104, cabinet 106, keyboard 108, and mouse 110. Mouse 110 may have one or more buttons such as mouse buttons 112. Cabinet 106 houses a CD-ROM drive 114, a system memory and a hard drive (see Fig. 33) which can be utilized to store and retrieve software programs incorporating code that implements the present invention, data for use with the present invention, and the like. Although a CD-ROM 116 is shown as an exemplary computer readable storage medium, other computer readable storage media including floppy disks, tape, flash memory, system memory, and hard drives may be utilized. Cabinet 106 also houses familiar computer components such as a central processor, system memory, hard disk, and the like.

Fig. 2 shows a system block diagram of computer system 100 that may be used to execute software embodiments of the present invention. As in Fig. 32, computer system 100 includes monitor 102 and keyboard 108. Computer system 100 further includes subsystems such as a central processor 102, system memory 120, I/O controller 122, display adapter 124, removable disk 126 (e.g., CD-ROM drive), fixed disk 128 (e.g., hard drive), network interface 130, and speaker 132. Other computer systems suitable for use with the present invention may include additional or fewer subsystems. For example, another computer system can include more than one processor 102 (i.e., a multi-processor system) or a cache memory.

Arrows such as 134 represent the system bus architecture of computer system 100. However, these arrows are illustrative of any interconnection scheme serving to link the subsystems. For example, a local bus can be utilized to connect the central processor to the system memory and display adapter. Computer system 100 shown in Fig. 33 is but an

example of a computer system suitable for use with the present invention.

The computer stores records relating to the polymorphisms of the record. Some such records record a polymorphism by reference to the position of a polymorphic site and the identity of base(s) occupying that site in one or more species. Some databases include records for at least ten polymorphic sites in at least ten of the sequences shown in Tables 1. Some databases include records for all of the polymorphic sites in at least one of the sequences shown in Tables 1. Some databases includes records for at least 100, 1000, or 2000 polymorphic sites shown in Tables 1. Some databases include records for all of the polymorphic sites shown in Tables 1.

The foregoing invention has been described in some detail by way of illustration and example, for purposes of clarity and understanding. It will be obvious to one of skill in the art that changes and modifications may be practiced within the scope of the appended claims. Therefore, it is to be understood that the above description is intended to be illustrative and not restrictive. The scope of the invention should, therefore, be determined not with reference to the above description, but should instead be determined with reference to the following appended claims, along with the full scope of equivalents to which such claims are entitled.

All patents, patent applications and publications cited in this application are hereby incorporated by reference in their entirety for all purposes to the same extent as if each individual patent, patent application or publication were so individually denoted.

WHAT IS CLAIMED IS:

1           1. A nucleic acid comprising at least 50  
2 contiguous bases from an rpoB sequence shown in Table 1.

1           2. The nucleic acid of claim 1 comprising a  
2 complete rpoB sequence shown in Table 1.

1           3. A set of probes perfectly complementary to and  
2 spanning a full-length sequence shown in Table 1.

1           4. A method of classifying a mycobacteria,  
2 comprising  
3           providing a sample comprising a mycobacterial rpoB  
4 target nucleic acid from a mycobacteria;  
5           determining the sequence of a segment of at least 50  
6 contiguous bases from the target nucleic acid;  
7           comparing the determined sequence to at least one  
8 sequence shown in Table 1;  
9           classifying the mycobacteria from the extent of  
10 similarity of the compared sequences.

1           5. The method of claim 4, wherein at least 100  
2 contiguous bases are determined from the target nucleic acid.

1           6. The method of claim 4, wherein the determined  
2 sequence is compared with at least ten sequences shown in  
3 Table 1.

1           7. A method of classifying a mycobacteria,  
2 comprising  
3           providing a sample comprising a mycobacterial rpoB  
4 target nucleic acid;  
5           determining the identity of one or more bases in the  
6 target sequence at one or more positions corresponding to one  
7 or more of the highlighted positions in a sequence shown in

8 Table 1, the identity of the one or more bases characterizing  
9 the species of mycobacteria that is present in the sample.

1 8. The method of claim 7, wherein the identity of  
2 at least 10 bases in the target nucleic acid at positions  
3 corresponding to highlighted positions in a sequence shown in  
4 Table 1 is determined.

1 9. The method of claim 8, wherein the identity of  
2 at least 20 bases in the target sequence at highlighted  
3 positions shown in Table 1 are identified.

1 10. The method of claim 9, further comprising  
2 comparing the 20 determined bases with 20 bases occupying  
3 corresponding positions in each of at least ten sequences from  
4 Table 1.

1 11. A sequence-specific polynucleotide probe or  
2 primer that hybridizes to a segment of a mycobacterial rpoB  
3 sequence shown in Table 1 or its complement without  
4 hybridizing to the M. tuberculosis sequence designated ATCC9-  
5 Mtb in Table 1 or its complement, wherein the segment includes  
6 a highlighted nucleotide position shown in Table 1.

1 10. The sequence-specific polynucleotide of claim 9  
2 that is probe.

1 11. The sequence-specific polynucleotide of  
2 claim 10, wherein a central position of the probe aligns with  
3 a highlighted nucleotide position shown in Tables 13.

1 12. The sequence-specific polynucleotide of claim 9  
2 that is a primer.



13. The sequence-specific polynucleotide of claim 12, wherein the 3' end of the primer aligns with a highlighted nucleotide position shown in Table 13.

14. The sequence-specific polynucleotide of claim 8 that is between 10 and 50 bases long.

19. A computer-readable storage medium for storing data for access by an application program being executed on a data processing system, comprising:

a data structure stored in the computer-readable storage medium, the data structure including information resident in a database used by the application program and including:

a plurality of records, each record comprising information identifying a polymorphism or sequence shown in Table 1.

20. The computer-readable storage medium of claim 19, wherein each record has a field identifying a base occupying a polymorphic site and a field identifying location of the polymorphic site.

21. The computer-readable storage medium of claim 19, wherein each record records a contiguous segment of at least 50 bases from an rpoB sequence shown in Table 1.

22. The computer-readable storage medium of claim 21, comprising at least ten records each recording a contiguous segment of at least 50 bases from at least ten rpoB sequences shown in Table 1.

**MYCOBACTERIAL *RpoB* SEQUENCES****ABSTRACT OF THE DISCLOSURE**

This invention provides polynucleotide probes, sequences and methods for speciating and phenotyping organisms, for example, using probes based on the *Mycobacterium tuberculosis rpoB* gene. The groups or species to which an organism belongs may be determined by comparing hybridization patterns of target nucleic acid from the organism to hybridization patterns in a database.

5

662040" 90258260

[illegible]



## Genchip0 Sequence Analysis

03/13/98 17:37:45

```

M77.txt      CC CAGGACGT GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M18.txt      GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M221.txt     GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M223.txt     GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M225.txt     GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M341.txt     GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M715.txt     GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M7470.txt    GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
Atcc4-go.txt GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
Atcc4-0-Gord.txt GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M78(1z).txt GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M79(1d).txt GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M80(1g).txt GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M81(1l).txt GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M82(1m).txt GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M83(mb).txt  GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M84(ow).txt GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M85(lb).txt  GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M86(1b).txt GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M87(mn).txt  GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M90(gordDB).txt GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
m126.txt     GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M128.txt     GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M103.txt     GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M7475.txt    GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M7476.txt    GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M930.txt     GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
Atcc5-int.txt GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
Atcc5-0int.txt GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M18.txt      GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M19.txt      GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M20.txt      GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M21.txt      GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M22.txt      GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M23.txt      GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M24.txt      GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M25.txt      GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M26.txt      GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M27.txt      GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M28.txt      GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M107.txt     GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M107.txt     GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M112.txt     GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M112.txt     GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
Atcc6-kan.txt GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
Atcc6-0kan.txt GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M1.txt       GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M2.txt       GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M3.txt       GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M4.txt       GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M5.txt       GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M6.txt       GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M7.txt       GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M9.txt       GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M57.txt      GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M58.txt      GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG
M59.txt      GGAAGGCGATCA CACCGGCA GACCGT GATCAACATCCGGCCCGTGTGGCCGG

```

[illegible]



[illegible]

[illegible]



[illegible]



147A (May 19) + x +

[illegible]

Atcc10-xen.txt



[illegible]

14



$$x + x^2 + x^3 + \dots + x^n + x^{n+1} + \dots$$



M74 (Mav12).txt

[illegible]

Atcc10-xen.txt

[illegible]

[illegible]

22



600.txt  
601.txt  
602.txt  
M106.txt  
M1216.txt  
M1218.txt  
M1226.txt  
M109.txt  
M111.txt  
M112.txt  
M113.txt  
M114.txt  
M1325.txt  
M1718.txt  
M1214.txt  
M1224.txt  
M1244.txt  
M1339.txt  
M1343.txt  
M1458.txt  
M1809.txt  
M1817.txt  
M1821.txt  
M1824.txt  
M1102.txt  
M1105.txt  
M1251.txt  
M1256.txt  
M1294.txt  
M1354.txt  
Atcc7-scr.txt  
Atcc7-0scr.txt  
M1121.txt  
M1249.txt  
M1372.txt  
M1378.txt  
M1484.txt  
M1556.txt  
M1563.txt  
M1586.txt  
Atcc8-sme.txt  
M35.txt  
M36.txt  
M37.txt  
M125.txt  
M1143.txt  
M1104.txt  
M1196.txt  
M1357.txt  
M1358.txt  
M1480.txt  
M1212.txt  
M1491.txt  
M1497.txt  
M1816.txt  
Atcc10-xen.txt



[illegible]

[illegible]

ATCC10-xonm.txt



[illegible]

03/13/98 17:37:45

M17.txt  
M18.txt  
MY221.txt  
MY223.txt  
MY225.txt  
MY341.txt  
MY715.txt  
MY470.txt  
ATCC4-g0.txt  
ATCC4-0-Gord.txt  
M78(1z).txt  
M79(jd).txt  
M80(1g).txt  
M81(1l).txt  
M82(1m).txt  
M83(mb).txt  
M84(ow).txt  
M85(1b).txt  
M86(zb).txt  
M87(wm).txt  
M90(gordDB).txt  
m126.txt  
M128.txt  
MY103.txt  
MY475.txt  
MY476.txt  
MY830.txt  
ATcc5-int.txt  
ATCC5-0int.txt  
M18.txt  
M19.txt  
M20.txt  
M21.txt  
M22.txt  
M23.txt  
M24.txt  
M25.txt  
M26.txt  
M27.txt  
M28.txt  
M107.txt  
MY107.txt  
MY112.txt  
MY312.txt  
ATcc6-kan.txt  
ATCC6-0kan.txt  
M1.txt  
M2.txt  
M3.txt  
M4.txt  
M6.txt  
M7.txt  
M9.txt  
M57.txt  
M58.txt  
M59.txt



[illegible][illegible]

03/13/98 17:37:45

[illegible]

ATCC9-Mtb.txt  
M621.txt  
Atcc91-av.txt  
M29.txt  
M30.txt  
M31.txt  
M32.txt  
M33.txt  
M34.txt  
M48-new.txt  
M49.txt  
M64 (Mav2).txt  
M65 (Mav3).txt  
M67 (Mav5).txt  
M69 (Mav7).txt  
M71 (Mav9).txt  
M91.txt  
M94.txt  
M95.txt  
M96.txt  
M100.txt  
M101.txt  
M102.txt  
M104.txt  
M105.txt  
M106.txt  
M111.txt  
M76.txt  
MY451.txt  
ATCC2-chelnew.txt  
M10.txt  
M11-662.txt  
M12.txt  
M13 2.662.txt  
M14.txt  
M15.txt  
M16.txt  
M17 2.662.txt  
M50.txt  
M51.txt  
M115.txt  
M116.txt  
MY109.txt  
MY119.txt  
MY200.txt  
MY207.txt  
M122.txt  
M123.txt  
M124.txt  
Atcc3-for.txt  
M53.txt  
M54.txt  
M55.txt  
M56.txt  
M74 (Mav12).txt

[illegible]

[illegible]

[illegible]

[illegible]

03/13/98 17:37:45

M17. txt  
M18. txt  
MY221. txt  
MY223. txt  
MY225. txt  
MY341. txt  
MY715. txt  
MY470. txt  
ATCC4-go. txt  
ATCC4-0-Gord. txt  
M78 (1z). txt  
M79 (j4d). txt  
M80 (1g). txt  
M81 (11). txt  
M82 (zm). txt  
M83 (mb). txt  
M84 (ow). txt  
M85 (1b). txt  
M86 (zb). txt  
M87 (wn). txt  
M90 (gordDB). txt  
m126. txt  
M128. txt  
MY103. txt  
MY475. txt  
MY476. txt  
MY830. txt  
ATcc5-int. txt  
ATCC5-0inc. txt  
M18. txt  
M19. txt  
M20. txt  
M21. txt  
M22. txt  
M23. txt  
M24. txt  
M25. txt  
M26. txt  
M27. txt  
M28. txt  
M107. txt  
MY107. txt  
MY112. txt  
MY312. txt  
ATcc6-kan. txt  
ATCC6-0Kan. txt  
M1. txt  
M2. txt  
M3. txt  
M4. txt  
M6. txt  
M7. txt  
M9. txt  
M57. txt  
M58. txt  
M59. txt  
M60. txt  
M61. txt  
M62. txt  
M63. txt  
M64. txt  
M65. txt  
M66. txt  
M67. txt  
M68. txt  
M69. txt  
M70. txt  
M71. txt  
M72. txt  
M73. txt  
M74. txt  
M75. txt  
M76. txt  
M77. txt  
M78. txt  
M79. txt  
M80. txt  
M81. txt  
M82. txt  
M83. txt  
M84. txt  
M85. txt  
M86. txt  
M87. txt  
M88. txt  
M89. txt  
M90. txt  
M91. txt  
M92. txt  
M93. txt  
M94. txt  
M95. txt  
M96. txt  
M97. txt  
M98. txt  
M99. txt  
M100. txt  
M101. txt  
M102. txt  
M103. txt  
M104. txt  
M105. txt  
M106. txt  
M107. txt  
M108. txt  
M109. txt  
M110. txt  
M111. txt  
M112. txt  
M113. txt  
M114. txt  
M115. txt  
M116. txt  
M117. txt  
M118. txt  
M119. txt  
M120. txt  
M121. txt  
M122. txt  
M123. txt  
M124. txt  
M125. txt  
M126. txt  
M127. txt  
M128. txt  
M129. txt  
M130. txt  
M131. txt  
M132. txt  
M133. txt  
M134. txt  
M135. txt  
M136. txt  
M137. txt  
M138. txt  
M139. txt  
M140. txt  
M141. txt  
M142. txt  
M143. txt  
M144. txt  
M145. txt  
M146. txt  
M147. txt  
M148. txt  
M149. txt  
M150. txt  
M151. txt  
M152. txt  
M153. txt  
M154. txt  
M155. txt  
M156. txt  
M157. txt  
M158. txt  
M159. txt  
M160. txt  
M161. txt  
M162. txt  
M163. txt  
M164. txt  
M165. txt  
M166. txt  
M167. txt  
M168. txt  
M169. txt  
M170. txt  
M171. txt  
M172. txt  
M173. txt  
M174. txt  
M175. txt  
M176. txt  
M177. txt  
M178. txt  
M179. txt  
M180. txt  
M181. txt  
M182. txt  
M183. txt  
M184. txt  
M185. txt  
M186. txt  
M187. txt  
M188. txt  
M189. txt  
M190. txt  
M191. txt  
M192. txt  
M193. txt  
M194. txt  
M195. txt  
M196. txt  
M197. txt  
M198. txt  
M199. txt  
M200. txt  
M201. txt  
M202. txt  
M203. txt  
M204. txt  
M205. txt  
M206. txt  
M207. txt  
M208. txt  
M209. txt  
M210. txt  
M211. txt  
M212. txt  
M213. txt  
M214. txt  
M215. txt  
M216. txt  
M217. txt  
M218. txt  
M219. txt  
M220. txt  
M221. txt  
M222. txt  
M223. txt  
M224. txt  
M225. txt  
M226. txt  
M227. txt  
M228. txt  
M229. txt  
M230. txt  
M231. txt  
M232. txt  
M233. txt  
M234. txt  
M235. txt  
M236. txt  
M237. txt  
M238. txt  
M239. txt  
M240. txt  
M241. txt  
M242. txt  
M243. txt  
M244. txt  
M245. txt  
M246. txt  
M247. txt  
M248. txt  
M249. txt  
M250. txt  
M251. txt  
M252. txt  
M253. txt  
M254. txt  
M255. txt  
M256. txt  
M257. txt  
M258. txt  
M259. txt  
M260. txt  
M261. txt  
M262. txt  
M263. txt  
M264. txt  
M265. txt  
M266. txt  
M267. txt  
M268. txt  
M269. txt  
M270. txt  
M271. txt  
M272. txt  
M273. txt  
M274. txt  
M275. txt  
M276. txt  
M277. txt  
M278. txt  
M279. txt  
M280. txt  
M281. txt  
M282. txt  
M283. txt  
M284. txt  
M285. txt  
M286. txt  
M287. txt  
M288. txt  
M289. txt  
M290. txt  
M291. txt  
M292. txt  
M293. txt  
M294. txt  
M295. txt  
M296. txt  
M297. txt  
M298. txt  
M299. txt  
M300. txt  
M301. txt  
M302. txt  
M303. txt  
M304. txt  
M305. txt  
M306. txt  
M307. txt  
M308. txt  
M309. txt  
M310. txt  
M311. txt  
M312. txt  
M313. txt  
M314. txt  
M315. txt  
M316. txt  
M317. txt  
M318. txt  
M319. txt  
M320. txt  
M321. txt  
M322. txt  
M323. txt  
M324. txt  
M325. txt  
M326. txt  
M327. txt  
M328. txt  
M329. txt  
M330. txt  
M331. txt  
M332. txt  
M333. txt  
M334. txt  
M335. txt  
M336. txt  
M337. txt  
M338. txt  
M339. txt  
M340. txt  
M341. txt  
M342. txt  
M343. txt  
M344. txt  
M345. txt  
M346. txt  
M347. txt  
M348. txt  
M349. txt  
M350. txt  
M351. txt  
M352. txt  
M353. txt  
M354. txt  
M355. txt  
M356. txt  
M357. txt  
M358. txt  
M359. txt  
M360. txt  
M361. txt  
M362. txt  
M363. txt  
M364. txt  
M365. txt  
M366. txt  
M367. txt  
M368. txt  
M369. txt  
M370. txt  
M371. txt  
M372. txt  
M373. txt  
M374. txt  
M375. txt  
M376. txt  
M377. txt  
M378. txt  
M379. txt  
M380. txt  
M381. txt  
M382. txt  
M383. txt  
M384. txt  
M385. txt  
M386. txt  
M387. txt  
M388. txt  
M389. txt  
M390. txt  
M391. txt  
M392. txt  
M393. txt  
M394. txt  
M395. txt  
M396. txt  
M397. txt  
M398. txt  
M399. txt  
M400. txt  
M401. txt  
M402. txt  
M403. txt  
M404. txt  
M405. txt  
M406. txt  
M407. txt  
M408. txt  
M409. txt  
M410. txt  
M411. txt  
M412. txt  
M413. txt  
M414. txt  
M415. txt  
M416. txt  
M417. txt  
M418. txt  
M419. txt  
M420. txt  
M421. txt  
M422. txt  
M423. txt  
M424. txt  
M425. txt  
M426. txt  
M427. txt  
M428. txt  
M429. txt  
M430. txt  
M431. txt  
M432. txt  
M433. txt  
M434. txt  
M435. txt  
M436. txt  
M437. txt  
M438. txt  
M439. txt  
M440. txt  
M441. txt  
M442. txt  
M443. txt  
M444. txt  
M445. txt  
M446. txt  
M447. txt  
M448. txt  
M449. txt  
M450. txt  
M451. txt  
M452. txt  
M453. txt  
M454. txt  
M455. txt  
M456. txt  
M457. txt  
M458. txt  
M459. txt  
M460. txt  
M461. txt  
M462. txt  
M463.



[illegible]

```
M38.txt
M39.txt
M40.txt
M41.txt
M42.txt
M43.txt
M44.txt
M45.txt
M46.txt
M47.txt
M68 (Mav6).txt
M89.txt
M66 (Mav4).txt
```

[illegible]

[illegible]

**C**

M17.txt  
M18.txt  
M121.txt  
MY223.txt  
MY225.txt  
My341.txt  
My715.txt  
MY470.txt  
Atcc4-go.txt  
ATCC4-0-Gord.txt  
M78(1z).txt  
M79(1d).txt  
M80(1g).txt  
M81(1l).txt  
M82(1m).txt  
M83(1nb).txt  
M84(1ow).txt  
M85(1p).txt  
M86(1zb).txt  
M87(1wn).txt  
M90(gordDB).txt  
m126.txt  
M128.txt  
My103.txt  
MY475.txt  
MY476.txt  
MY830.txt  
Atcc5-Int.txt  
ATCC5-0Int.txt  
M18.txt  
M19.txt  
M20.txt  
M21.txt  
M22.txt  
M23.txt  
M24.txt  
M25.txt  
M26.txt  
M27.txt  
M28.txt  
M107.txt  
MY107.txt  
MY112.txt  
MY312.txt  
Atcc6-Kan.txt  
ATCC6-0Kan.txt  
M1.txt  
M2.txt  
M3.txt  
M4.txt  
M6.txt  
M7.txt  
M9.txt  
M57.txt  
M58.txt  
M59.txt



[illegible]

Year	1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100
1990	1991	1992	1993	1994	1995	1996	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012	2013	2014	2015	2016	2017	2018	2019	2020	2021	2022	2023	2024	2025	2026	2027	2028	2029	2030	2031	2032	2033	2034	2035	2036	2037	2038	2039	2040	2041	2042	2043	2044	2045	2046	2047	2048	2049	2050	2051	2052	2053	2054	2055	2056	2057	2058	2059	2060	2061	2062	2063	2064	2065	2066	2067	2068	2069	2070	2071	2072	2073	2074	2075	2076	2077	2078	2079	2080	2081	2082	2083	2084	2085	2086	2087	2088	2089	2090	2091	2092	2093	2094	2095	2096	2097	2098	2099	2100	

[illegible]

46





M77.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M118.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
MY221.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
MY223.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
MY225.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
MY341.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
MY715.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
MY470.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
Atcc4-go.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
ATCC4-0-Gord.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
MY78(1z).txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
MY9(1d).txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M80(1g).txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M81(1l).txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M82(1m).txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M83(mb).txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M84(ow).txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M85(1b).txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M86(1d).txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M87(wt).txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M90(gordDB).txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
m126.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M128.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
MY103.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
MY475.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
MY476.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
MY830.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
Atcc5-1nt.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
ATCC5-0int.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M18.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M19.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M20.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M21.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M22.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M23.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M24.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M25.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M26.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M27.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M28.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M107.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
MY107.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
MY112.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
MY312.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
Atcc6-kan.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
ATCC6-0kan.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M1.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M2.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M3.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M4.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M6.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M7.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M9.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M57.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
M58.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT
MEM.txt	AGGCCCCCGCTGGTGGG	ATGGAGCTGCGCGCGCGGCGATCGACCGCGGCGACGT

45

[illegible]

Affy#	SAMPLE ID#	Alt. ID	SPECIES
MY621		ATCC	<i>M. abscessus</i>
ATCC1	25291	ATCC-av	<i>M. avium</i>
M100	60300	MAC	<i>M. avium</i>
M101	60112	MAC	<i>M. avium</i>
M102	60268	MAC	<i>M. avium</i>
M103	60270	MAC	<i>M. avium</i>
M104	60272	MAC	<i>M. avium</i>
M105	60293	MAC	<i>M. avium</i>
M106	60313	MAC	<i>M. avium</i>
M107	60345	MAC	<i>M. avium</i>
M29	95-1764		<i>M. avium</i>
M30	95-1766		<i>M. avium</i>
M31	95-1768		<i>M. avium</i>
M32	95-1770		<i>M. avium</i>
M33	95-1775		<i>M. avium</i>
M34	95-1776		<i>M. avium</i>
M48	95-1765		<i>M. avium</i>
M49	95-1769		<i>M. avium</i>
M63	MAC #1	MAC	<i>M. avium</i>
M64	MAC #2	MAC	<i>M. avium</i>
M65	MAC #3	MAC	<i>M. avium</i>
M67	MAC #5	MAC	<i>M. avium</i>
M69	MAC #7	MAC	<i>M. avium</i>
M70	MAC #8	MAC	<i>M. avium</i>
M71	MAC #9	MAC	<i>M. avium</i>
M72	MAC #10	MAC	<i>M. avium</i>
M91	FM	<i>avium-intracell. FM(MAC)</i>	<i>M. avium</i>
M92	60040	MAC	<i>M. avium</i>
M93	60042	MAC	<i>M. avium</i>
M94	60049	MAC	<i>M. avium</i>
M95	60051	MAC	<i>M. avium</i>
M96	60110	MAC	<i>M. avium</i>
M97	60116	MAC	<i>M. avium</i>
M98	60123	MAC	<i>M. avium</i>
M99	60176	MAC	<i>M. avium</i>
M76	92-773		<i>M. bovis</i>
MY451			<i>M. bovis</i>
ATCC2	35752	ATCC-chel(new)	<i>M. chelonae</i>
M10	95A9151		<i>M. chelonae</i>
M11	95A0477		<i>M. chelonae</i>
M115	60121		<i>M. chelonae</i>
M116	52942		<i>M. chelonae</i>
M117	43192		<i>M. chelonae</i>
M118	53180		<i>M. chelonae</i>
M119	53131		<i>M. chelonae</i>
M12	95A4883		<i>M. chelonae</i>
M120	52923		<i>M. chelonae</i>
M121	52919		<i>M. chelonae</i>
M13	95A2611		<i>M. chelonae</i>
M14	95A0779		<i>M. chelonae</i>
M15	95A8654		<i>M. chelonae</i>
M16	95A8882		<i>M. chelonae</i>
M17	95A8881		<i>M. chelonae</i>
M50	95A11814		<i>M. chelonae</i>
M51	95A1102		<i>M. chelonae</i>
M75	#13	MAC#13	<i>M. chelonae</i>
MY109			<i>M. chelonae</i>
MY200			<i>M. chelonae</i>

FIG. 2 (2 of 6)

MY207			<i>M. chelonae</i>
MY209			<i>M. chelonae</i>
M122	60025		<i>M. flavescens</i>
M123	60078		<i>M. flavescens</i>
M124	60252		<i>M. flavescens</i>
ATCC3	6841	ATCC-for	<i>M. fortuitum</i>
M53	60305		<i>M. fortuitum</i>
M54	60344		<i>M. fortuitum</i>
M55	60435		<i>M. fortuitum</i>
M56	60447		<i>M. fortuitum</i>
M74	#12	MAC#12	<i>M. fortuitum</i>
M88	CH	fort. CH	<i>M. fortuitum</i>
MY221			<i>M. fortuitum</i>
MY223			<i>M. fortuitum</i>
MY225			<i>M. fortuitum</i>
MY341			<i>M. fortuitum</i>
MY715			<i>M. fortuitum</i>
MY470			<i>M. genevieve</i>
ATCC4	14470	ATCC-go	<i>M. gordonae</i>
ATCC4-0		gord.	<i>M. gordonae</i>
M125	60068		<i>M. gordonae</i>
M126	60182		<i>M. gordonae</i>
M127	60214		<i>M. gordonae</i>
M128	60283		<i>M. gordonae</i>
M78	92-942	gord. LZ	<i>M. gordonae</i>
M79	93-692	gord. JD	<i>M. gordonae</i>
M80	94-84	gord. LG	<i>M. gordonae</i>
M81	93-1231	gord. LL	<i>M. gordonae</i>
M82	93-483	gord. RM	<i>M. gordonae</i>
M83	92-1219	gord. MB	<i>M. gordonae</i>
M84	91-1131	gord. OW	<i>M. gordonae</i>
M85	91-1478	gord. LB	<i>M. gordonae</i>
M86	92-642	gord. RB	<i>M. gordonae</i>
M87	93-1180	gord. WN	<i>M. gordonae</i>
M90	DB	gord. DB	<i>M. gordonae</i>
MY103			<i>M. gordonae</i>
MY475			<i>M. gordonae</i>
MY476			<i>M. gordonae</i>
MY746			<i>M. gordonae</i>
MY830			<i>M. gordonae</i>
ATCC5		ATCC-int	<i>M. intracellulare</i>
ATCC5-0		intra	<i>M. intracellulare</i>
M18	95-1778		<i>M. intracellulare</i>
M19	95-1780		<i>M. intracellulare</i>
M20	95-1781		<i>M. intracellulare</i>
M21	95-1782		<i>M. intracellulare</i>
M22	95-1790		<i>M. intracellulare</i>
M23	95-1794		<i>M. intracellulare</i>
M24	95-1796		<i>M. intracellulare</i>
M25	95-1777		<i>M. intracellulare</i>
M26	95-1779		<i>M. intracellulare</i>
M27	95-1760		<i>M. intracellulare</i>
M28	95-1761		<i>M. intracellulare</i>
ATCC6	12478	ATCC-kan	<i>M. kansasii</i>
ATCC6-0		kans.	<i>M. kansasii</i>
M1	95A5375		<i>M. kansasii</i>
M2	95A10299		<i>M. kansasii</i>
M3	96A0020		<i>M. kansasii</i>
M4	95A3977		<i>M. kansasii</i>

662040" 20233260

FIG. 2 (3 of 6)

M5	95A4739		<i>M. kansasii</i>
M52	95A5381		<i>M. kansasii</i>
M57	60163		<i>M. kansasii</i>
M58	60180		<i>M. kansasii</i>
M59	60207		<i>M. kansasii</i>
M6	95A2695		<i>M. kansasii</i>
M60	60294		<i>M. kansasii</i>
M61	60308		<i>M. kansasii</i>
M62	60314		<i>M. kansasii</i>
M7	95A2694		<i>M. kansasii</i>
M73	#11	MAC#11	<i>M. kansasii</i>
M8	94A9042		<i>M. kansasii</i>
M9	95A1275		<i>M. kansasii</i>
MY106			<i>M. kansasii</i>
MY141			<i>M. kansasii</i>
MY216			<i>M. kansasii</i>
MY218			<i>M. kansasii</i>
MY226			<i>M. kansasii</i>
M108	60044		<i>M. malmoense</i>
M109	60149		<i>M. malmoense</i>
M110	60211		<i>M. malmoense</i>
M111	60202		<i>M. malmoense</i>
M112	60085		<i>M. malmoense</i>
M113	60047		<i>M. malmoense</i>
M114	60185		<i>M. malmoense</i>
MY325		ATCC	<i>M. malmoense</i>
MY718		malmo	<i>M. malmoense</i>
MY214			<i>M. marinum</i>
MY224			<i>M. marinum</i>
MY244			<i>M. marinum</i>
MY339			<i>M. marinum</i>
MY343			<i>M. marinum</i>
MY458		ATCC	<i>M. mucogenicum</i>
MY809			<i>M. mucogenicum</i>
MY817			<i>M. mucogenicum</i>
MY821			<i>M. mucogenicum</i>
MY824			<i>M. mucogenicum</i>
MY102			<i>M. nonchromogenicum</i>
MY105			<i>M. nonchromogenicum</i>
MY251			<i>M. nonchromogenicum</i>
MY256			<i>M. nonchromogenicum</i>
MY294			<i>M. nonchromogenicum</i>
ATCC7	19981	ATCC-scr	<i>M. scrofulaceum</i>
ATCC7-0		scrof.	<i>M. scrofulaceum</i>
MY121			<i>M. scrofulaceum</i>
MY249			<i>M. scrofulaceum</i>
MY372			<i>M. scrofulaceum</i>
MY378			<i>M. scrofulaceum</i>
MY484			<i>M. simiae</i>
MY556			<i>M. simiae</i>
MY563			<i>M. simiae</i>
MY586			<i>M. simiae</i>
ATCC8	19420	ATCC-smc	<i>M. smegmatis</i>
M35	95A1072		<i>M. smegmatis</i>
M36	95A8183		<i>M. smegmatis</i>
M37	95A4990		<i>M. smegmatis</i>
M77	92-144	smeg. JL	<i>M. smegmatis</i>
MY143		ATCC	<i>M. smegmatis</i>
MY104			<i>M. szulgai</i>

662040" 20252260

FIG 2 (4 of 6)

MY196			<i>M. szulgai</i>
MY357			<i>M. szulgai</i>
MY358			<i>M. szulgai</i>
MY480			<i>M. szulgai</i>
TB74	C.17.96.5		<i>M. lab M160 DR</i>
MY387			<i>M. tb</i>
MY418			<i>M. tb</i>
MY437			<i>M. tb</i>
MY462			<i>M. tb</i>
TB59	C.18.96.1		<i>M. tb H37rv DR</i>
TB67	C.18.96.1		<i>M. tb H37rv DR</i>
TB73	C.17.96.1		<i>M. tb H37rv DR</i>
TB60	C.18.96.2		<i>M. tb J35 DR</i>
TB65	C.22.96.9		<i>M. tb M101 DR</i>
TB62	C.18.96.4		<i>M. tb M104 DR</i>
TB69	C.18.96.3		<i>M. tb M104 DR</i>
TB72	C.18.96.7		<i>M. tb M104DR</i>
TB66	C.22.96.10		<i>M. tb M112 DR</i>
TB63	C.18.96.5		<i>M. tb M140 DR</i>
TB64	C.18.96.6		<i>M. tb M160 DR</i>
TB70	C.18.96.4		<i>M. tb M160 DR</i>
TB61	C.18.96.3		<i>M. tb M60 DR</i>
TB68	C.18.96.2		<i>M. tb M60 DR</i>
TB71	C.18.96.6		<i>M. tb M60 DR</i>
MY212			<i>M. terrae</i>
MY354			<i>M. terrae</i>
MY491			<i>M. terrae</i>
MY497			<i>M. terrae</i>
MY816			<i>M. triplex</i>
ATCC9	27294	Mtb	<i>M. tuberculosis</i>
ATCC9-0		TB2020	<i>M. tuberculosis</i>
N/A	93-1071		<i>M. tuberculosis</i>
N/A	93-336		<i>M. tuberculosis</i>
N/A	92-852		<i>M. tuberculosis</i>
N/A	92-1005		<i>M. tuberculosis</i>
N/A	92-243		<i>M. tuberculosis</i>
N/A	92-304		<i>M. tuberculosis</i>
N/A	92-199		<i>M. tuberculosis</i>
N/A	92-197		<i>M. tuberculosis</i>
N/A	92-484		<i>M. tuberculosis</i>
N/A	94-577		<i>M. tuberculosis</i>
TB1	936		<i>M. tuberculosis</i>
TB10	1122		<i>M. tuberculosis</i>
TB11	3407		<i>M. tuberculosis</i>
TB12	978		<i>M. tuberculosis</i>
TB13	3553		<i>M. tuberculosis</i>
TB14	3466		<i>M. tuberculosis</i>
TB15	2163		<i>M. tuberculosis</i>
TB16	DW	DW	<i>M. tuberculosis</i>
TB17	CB	CB	<i>M. tuberculosis</i>
TB18	PB	PB	<i>M. tuberculosis</i>
TB19	AA	AA	<i>M. tuberculosis</i>
TB2	M0404A		<i>M. tuberculosis</i>
TB20	3492		<i>M. tuberculosis</i>
TB21	1435		<i>M. tuberculosis</i>
TB22	696		<i>M. tuberculosis</i>
TB23	2268		<i>M. tuberculosis</i>
TB24	3455		<i>M. tuberculosis</i>
TB25	37		<i>M. tuberculosis</i>

FIG. 2 (5 of 6)

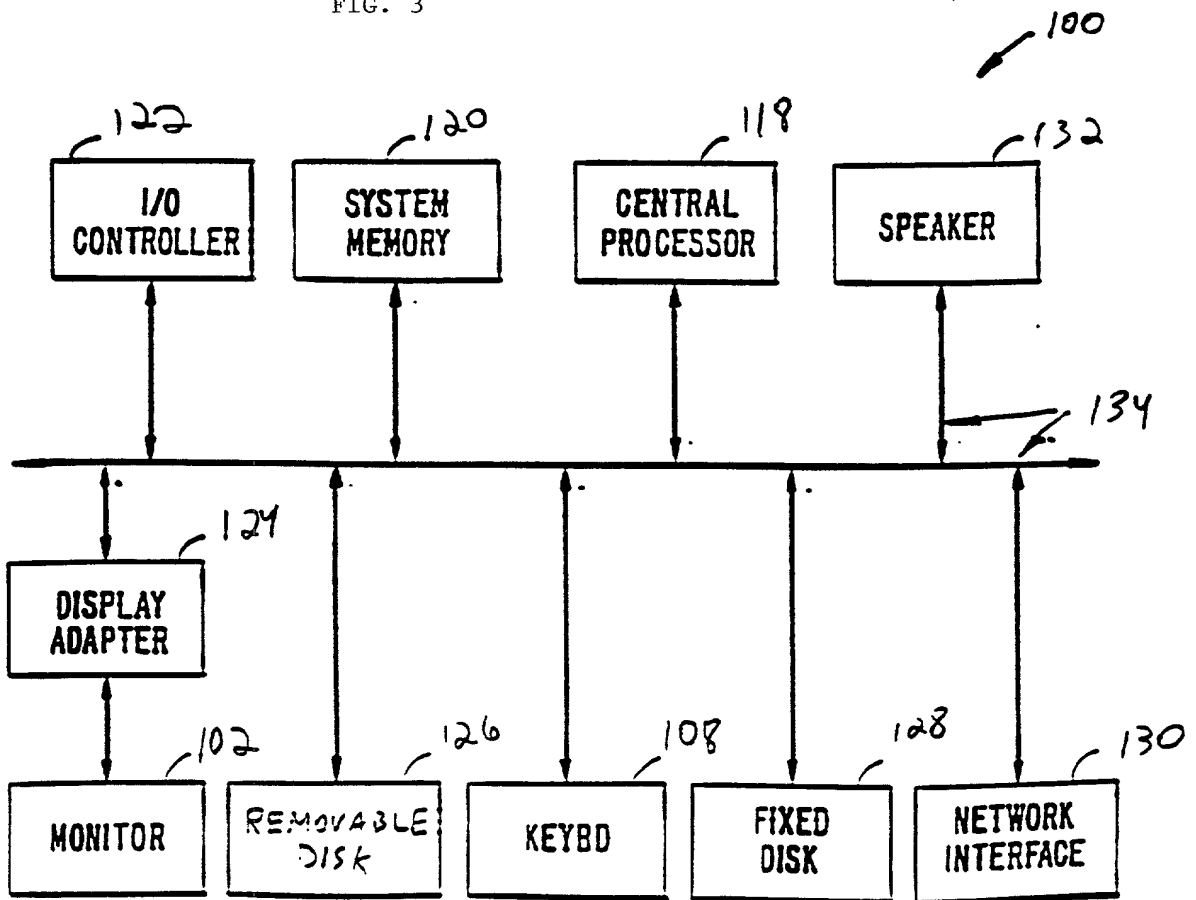
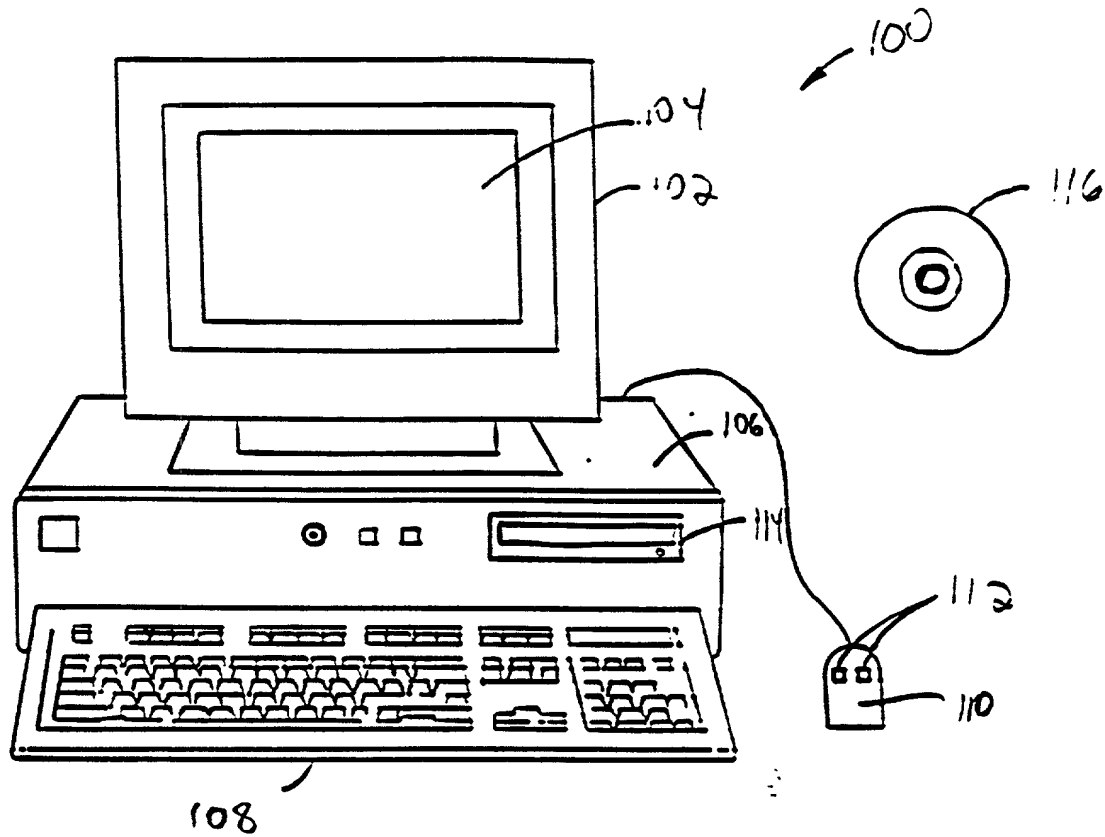
TB26	173		<i>M. tuberculosis</i>
TB27	230		<i>M. tuberculosis</i>
TB28	2519		<i>M. tuberculosis</i>
TB29	T29233		<i>M. tuberculosis</i>
TB3	1231		<i>M. tuberculosis</i>
TB30	SP	SP	<i>M. tuberculosis</i>
TB31	3201		<i>M. tuberculosis</i>
TB32	3219		<i>M. tuberculosis</i>
TB33	80		<i>M. tuberculosis</i>
TB34	3442		<i>M. tuberculosis</i>
TB35	3502		<i>M. tuberculosis</i>
TB36	3759		<i>M. tuberculosis</i>
TB37	1295		<i>M. tuberculosis</i>
TB38	337		<i>M. tuberculosis</i>
TB39	394		<i>M. tuberculosis</i>
TB4	914		<i>M. tuberculosis</i>
TB40	489		<i>M. tuberculosis</i>
TB41	535		<i>M. tuberculosis</i>
TB42	607		<i>M. tuberculosis</i>
TB43	707		<i>M. tuberculosis</i>
TB44	692		<i>M. tuberculosis</i>
TB45	2408		<i>M. tuberculosis</i>
TB46	1069		<i>M. tuberculosis</i>
TB47	M3262A		<i>M. tuberculosis</i>
TB48	1336		<i>M. tuberculosis</i>
TB49	1368		<i>M. tuberculosis</i>
TB5	1145		<i>M. tuberculosis</i>
TB50	65		<i>M. tuberculosis</i>
TB51	727		<i>M. tuberculosis</i>
TB52	3455		<i>M. tuberculosis</i>
TB53	3506		<i>M. tuberculosis</i>
TB54	9600367		<i>M. tuberculosis</i>
TB55	9600173		<i>M. tuberculosis</i>
TB56	9503471		<i>M. tuberculosis</i>
TB57	9600309		<i>M. tuberculosis</i>
TB58	9600230		<i>M. tuberculosis</i>
TB6	1417		<i>M. tuberculosis</i>
TB7	SM2341		<i>M. tuberculosis</i>
TB75	2098		<i>M. tuberculosis</i>
TB76	173/1		<i>M. tuberculosis</i>
TB77	1122/1		<i>M. tuberculosis</i>
TB78	1417/1		<i>M. tuberculosis</i>
TB8	1587		<i>M. tuberculosis</i>
TB9	M7032A		<i>M. tuberculosis</i>
ATCC10	19250	ATCC-xen	<i>M. xenopi</i>
M129	60133		<i>M. xenopi</i>
M130	60200		<i>M. xenopi</i>
M131	60365		<i>M. xenopi</i>
M132	60387		<i>M. xenopi</i>
M38	95A5208		<i>M. xenopi</i>
M39	95A5399		<i>M. xenopi</i>
M40	95A3938		<i>M. xenopi</i>
M41	95A8762		<i>M. xenopi</i>
M42	95A0933		<i>M. xenopi</i>
M43	95A4320		<i>M. xenopi</i>
M44	95A3478		<i>M. xenopi</i>
M45	95A2997		<i>M. xenopi</i>
M46	95A8383		<i>M. xenopi</i>
M47	95A4319		<i>M. xenopi</i>



FIG. 2 (6 of 6)

M68	MAC #8	MAC #8	<i>M. xenopi</i>
M89	SG	xen. SG	<i>M. xenopi</i>
MY219			<i>M. xenopi</i>
MY250			<i>M. xenopi</i>
MY252			<i>M. xenopi</i>
MY254			<i>M. xenopi</i>
MY255			<i>M. xenopi</i>
MY107			MAC
MY111			MAC
MY112			MAC
MY312			MAC
M68	MAC #4	MAC #4	unique

662040" 90252260



**DECLARATION AND POWER OF ATTORNEY**

As a below named inventor, I declare that:

My residence, post office address and citizenship are as stated below next to my name; I believe I am the original, first and sole inventor (if only one name is listed below) or an original, first and joint inventor (if plural inventors are named below) of the subject matter which is claimed and for which a patent is sought on the invention entitled: MYCOBACTERIAL RpoB SEQUENCES the specification of which X is attached hereto or \_\_\_\_ was filed on \_\_\_\_\_ as Application No. \_\_\_\_\_ and was amended on \_\_\_\_\_ (if applicable).

I have reviewed and understand the contents of the above identified specification, including the claims, as amended by any amendment referred to above. I acknowledge the duty to disclose information which is material to the examination of this application in accordance with Title 37, Code of Federal Regulations, Section 1.56. I claim foreign priority benefits under Title 35, United States Code, Section 119 of any foreign application(s) for patent or inventor's certificate listed below and have also identified below any foreign application for patent or inventor's certificate having a filing date before that of the application on which priority is claimed.

**Prior Foreign Application(s)**

Country	Application No.	Date of Filing	Priority Claimed Under 35 USC 119

Thereby claim the benefit under Title 35, United States Code § 119(e) of any United States provisional application(s) listed below:

Application No.	Filing Date
60/080,616	April 3, 1998

I claim the benefit under Title 35, United States Code, Section 120 of any United States application(s) listed below and, insofar as the subject matter of each of the claims of this application is not disclosed in the prior United States application in the manner provided by the first paragraph of Title 35, United States Code, Section 112, I acknowledge the duty to disclose material information as defined in Title 37, Code of Federal Regulations, Section 1.56 which occurred between the filing date of the prior application and the national or PCT international filing date of this application:

Application No.	Date of Filing	Status

**POWER OF ATTORNEY:** As a named inventor, I hereby appoint the following attorney(s) and/or agent(s) to prosecute this application and transact all business in the Patent and Trademark Office connected therewith.

Joe Liebeschuetz, Reg. No. 37,505  
Philip L. McGarrigle, Reg. No. 31,395  
William M. Smith, Reg. No. 30,223

Send Correspondence to: <b>Joe Liebeschuetz</b> <b>TOWNSEND and TOWNSEND and CREW LLP</b> <b>Two Embarcadero Center, 8<sup>th</sup> Floor</b> <b>San Francisco, California 94111-3834</b>	Direct Telephone Calls to: (Name, Reg. No., Telephone No.) Name: Joe Liebeschuetz Reg. No.: 37,505 Telephone: 650-326-2400
---	--

Full Name of Inventor 1:	Last Name: <b>Gingeras</b>	First Name: <b>Thomas</b>	Middle Name or Initial:	
Residence & Citizenship:	City:	State/Foreign Country:	Country of Citizenship:	
Post Office Address:	Post Office Address:	City:	State/Country:	Postal Code:
Full Name of Inventor 2:	Last Name: <b>Dreukow</b>	First Name: <b>Jorg</b>	Middle Name or Initial:	
Residence & Citizenship:	City:	State/Foreign Country:	Country of Citizenship:	
Post Office Address:	Post Office Address:	City:	State/Country:	Postal Code:

I further declare that all statements made herein of my own knowledge are true and that all statements made on information and belief are believed to be true; and further that these statements were made with the knowledge that willful false statements and the like so made are punishable by fine or imprisonment, or both, under Section 1001 of Title 18 of the United States Code, and that such willful false statements may jeopardize the validity of the application or any patent issuing thereon.

Signature of Inventor 1 Thomas Gingeras  _____	Signature of Inventor 2 Jorg(e) Dreukow  _____	Signature of Inventor 3  _____
Date	Date	Date

184402 v1

184402 v1